

Parameter Estimation in Two-type Continuous-state Branching Processes with Immigration

Wei Xu

School of Mathematical Sciences, Beijing Normal University

Beijing 100875, People's Republic of China

E-mails: xuwei@mail.bnu.edu.cn *or* xuweijlu.2008@163.com

Abstract: We study the estimation of two-type continuous-state branching processes with immigration (CBI-processes). The ergodicity of the processes is proved. We also establish the strong consistency and central limit theorems of the conditional least squares estimators and the weighted conditional least squares estimators of the drift and diffusion coefficients based on low frequency observations.

Key words and phrases: Two-type, continuous-state branching process with immigration, stochastic differential equation, conditional least squares estimator, weighted conditional least squares estimator, consistency, central limit theorem

1 Introduction

Branching processes have been used widely not only in biology, but also in financial world. For example, Galton-Watson branching processes with immigration (GWI-processes) are used to study the evolution of different species. Continuous-state branching processes (CB-processes) were first introduced by Jiřina (1958). In particular, a continuous CB-process can be obtained as the unique solution of a stochastic equation system driven by Brownian motion. Kawazu and Watanabe (1971) constructed continuous-state branching processes with immigration (CBI-processes). In view of the results of Dawson and Li (2006), a general single-type CBI-process is the unique strong solution of a stochastic equation driven by Brownian motions and Poisson random measures. The two-type CB-processes was first be introduced by Watanabe (1969). Ma (2012) proved the existence and uniqueness of the strong solution of a two-dimensional stochastic integral equation system with jumps. He also showed that the unique solution is a two-type CBI-process. In financial world, multitype CBI-processes are used to describe the relations of the prices of different assets and interest rates of different currencies.

Firstly, we introduce a special continuous single-type CBI-process defined by the following equation:

$$dX_t = (a - bX_t)dt + \sigma\sqrt{X_t}dB_t, \quad (1.1)$$

where $(a, b, \sigma) \in (0, +\infty)^3$ and B_t is a standard Brownian motion. In fact, the solution of (1.1) is also known as the *Cox-Ingersoll-Ross model* (CIR-model) introduced by Cox et al. (1985) for the term structure of interest rates. The above equation was also studied in Ikeda and Watanabe (1989) and Revuz and Yor (1991). The basic theory of general CBI-processes was developed in Li (2011). The appealing properties of this process are as follows:

- (1) The process stays nonnegative.
- (2) It converges to a steady-state law with mean a/b , the so-called long-term value, with speed of adjustment b .
- (3) The incremental variance is proportional to its current value.

However, the one-dimensional CIR-model doesn't describe the connection among interest rates of different currencies. In order to give more objective description to the financial environment, we need to deal with

multifactor CIR-model or the multitype CBI-processes. In order to make the presentation easier, we just analyze the two-type CBI-processes defined by the following equation:

$$\begin{cases} dX_1(t) = (a_1 - b_{11}X_1(t) + b_{12}X_2(t))dt + \sigma_1\sqrt{X_1(t)}dB_1(t), \\ dX_2(t) = (a_2 + b_{21}X_1(t) - b_{22}X_2(t))dt + \sigma_2\sqrt{X_2(t)}dB_2(t), \end{cases} \quad (1.2)$$

where $\theta = (a_1, a_2, b_{11}, b_{12}, b_{21}, b_{22}, \sigma_1, \sigma_2) \in (0, +\infty)^3 \times [0, \infty)^2 \times (0, \infty)^3$, and $B_i(t), i = 1, 2$ are independent standard Brownian motions. The existence and uniqueness of the solution to (1.2) were proved in Ma (2012). We can rewrite (1.2) into the vector form:

$$dX_t = (A - BX_t)dt + \Sigma\sqrt{X_t}dW_t, \quad (1.3)$$

where

$$X_t = \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix}, \quad A = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & -b_{12} \\ -b_{21} & b_{22} \end{pmatrix},$$

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \quad \sqrt{X_t} = \begin{pmatrix} \sqrt{X_1(t)} & 0 \\ 0 & \sqrt{X_2(t)} \end{pmatrix}, \quad W_t = \begin{pmatrix} B_1(t) \\ B_2(t) \end{pmatrix}.$$

We can use a special form of (1.2) to describe the relations among the interest rates of different currencies. In currency market, we assume $X_1(t)$ is the interest rate of a very strong and influential currency, and $X_2(t)$ is the interest rate of a less influential currency. The situation can be described by the following stochastic equation:

$$\begin{cases} dX_1(t) = b_{11}(\frac{a_1}{b_{11}} - X_1(t))dt + \sigma_1\sqrt{X_1(t)}dB_1(t), \\ dX_2(t) = b_{22}(\frac{a_2}{b_{22}} + \frac{b_{21}}{b_{22}}X_1(t) - X_2(t))dt + \sigma_2\sqrt{X_2(t)}dB_2(t). \end{cases}$$

Here, the first equation gives the evolution of $X_1(t)$, which is just a one-dimensional CIR- model. The second equation describes the evolution of $X_2(t)$, which is affected not only by the random noise, but also by $X_1(t)$. Specifically, the second coordination $X_2(t)$ has the following properties:

- (1) It stays nonnegative.
- (2) It converges to a steady-state law with mean $a_2/b_{22} + (b_{21}/b_{22})(a_1/b_{11})$ (this can be easily got from (3.1) with $t \rightarrow \infty$), the so-called long-term value, where the second term is contributed by $X_1(t)$.
- (3) Its incremental variance is proportional to its current value.

If b_{12} and b_{21} are neither zero, then (1.2) can account for the fluctuations of the rate of two currencies that affect with each other. Furthermore, (1.2) can also be used to analyze the corporate profitability and the market yield. For example,

- (1) In a perfectly competitive market, we can use (1.2) with $b_{12} = 0$ to analyze the relationship between the corporate profitability and the yield of its corresponding industry.
- (2) In an oligopoly market, we can use (1.2) to describe the relationship among profitability of the different enterprises.
- (3) However, in pure monopoly market, the one-dimensional CIR-model is suitable enough to analyze the corporate profitability or the market yield.

However, before using (1.3) to solve the practical problems, we need to estimate the parameters in the equation based on the historical information. For the single-type CBI-processes, the approaches to parameter estimation can be found in Longstaff and Schwartz (1992) and Bibby and Sørensen (1995). Overbeck and Rydén (1997) also gave the *conditional least squares estimators* (CLSEs). Estimators of the matrix of offspring means and the vector of stationary means in a multitype GWI-process had been given in Quine and Durham (1977). For multitype GWI-process, the *weighted conditional least squares*

estimator (WCLSE) of the mean matrix was developed in Shete and Sriram (2003). The asymptotically properties of CLSEs of GWI-processes with general offspring laws were studied in Venkataraman (1982) and Wei and Winnicki (1989). The asymptotics of CLSEs and WCLSEs of a stable CIR-model was studied in Li and Ma (2013). It is well-known that the CBI-processes are special examples of the affine Markov processes studied in Duffie et al. (2003). The ergodicity and estimation of some different two-dimensional affine processes were studied in Barczy et al. (2013a, 2013b, 2013c).

In this work, we give the CLSEs and the WCLSEs of the parameters in (1.3) using low frequency observations at equidistant time points $\{k\Delta : k = 0, 1, \dots, n\}$ of a single realization $\{X_t : t \geq 0\}$, where $X_t = (X_1(t), X_2(t))^T$. For simplicity, we take $\Delta = 1$, but all the results presented below can be modified to the general case. This is based on the minimization of a sum of squared deviation about conditional expectations developed in Klimko and Nelson (1978), who applied their results to the CLSEs of the offspring and immigration means of subcritical GWI-processes. Then, as Overbeck and Rydén (1997), we shall study the consistency and the central limit theorems of CLSEs and WCLSEs.

The paper is organized as follows. In Section 2, we give the ergodicity of the two-type CBI-process, which is essentially necessary for the study of the estimators. Section 3 is devoted to the study of the CLSEs and WCLSEs of (A, B) and Σ . The consistency and asymptotic normality of the CLSEs and WCLSEs are given in the Section 4. All the proofs are given in Section 5.

2 Multitype CBI-processes and ergodicity

In this section, firstly we give the definition and a few properties of two-type CBI-processes. In particular, we show that the solution of (1.3) is a two-type CBI-process. Secondly, we show that the two-type CBI-process is ergodic under a weak condition. These results are very important to study the consistency and asymptotic normality of the estimators.

Let $\mathbf{D} = [0, \infty)^2$. In view of (1.3), we consider the branching mechanisms $\phi_i, i = 1, 2$, with representation:

$$\begin{cases} \phi_1(\lambda) = b_{11}\lambda_1 - b_{12}\lambda_2 + \frac{\sigma_1^2}{2}\lambda_1^2, \\ \phi_2(\lambda) = -b_{21}\lambda_1 + b_{22}\lambda_2 + \frac{\sigma_2^2}{2}\lambda_2^2, \end{cases} \quad (2.1)$$

where $\lambda = (\lambda_1, \lambda_2)^T \in \mathbf{D}$. Next we give the definition of the two-type CBI-processes.

Definition 2.1 A Markov process $X_t = (X_1(t), X_2(t))$ in \mathbf{D} is called a two-type continuous-state branching process with immigration (CBI-process), if it has transition semigroup $(Q_t)_{t \geq 0}$ given by

$$\int_{\mathbf{D}} e^{-\langle y, \lambda \rangle} Q_t(x, dy) = \exp \left\{ -\langle x, v_t(\lambda) \rangle - \int_0^t \langle A, v_s(\lambda) \rangle ds \right\}, \quad \lambda, x \in \mathbf{D}, \quad (2.2)$$

where A is given in (1.3) and $v_t(\lambda) = (v_1(t, \lambda), v_2(t, \lambda))^T$ is the unique solution of

$$\begin{cases} \frac{\partial}{\partial t} v_1(t, \lambda) = -\phi_1(v_t(\lambda)) = -b_{11}v_1(t, \lambda) + b_{12}v_2(t, \lambda) - \frac{\sigma_1^2}{2}v_1(t, \lambda)^2, \\ \frac{\partial}{\partial t} v_2(t, \lambda) = -\phi_2(v_t(\lambda)) = b_{21}v_1(t, \lambda) - b_{22}v_2(t, \lambda) - \frac{\sigma_2^2}{2}v_2(t, \lambda)^2, \end{cases} \quad (2.3)$$

with the initial condition $v_0(\lambda) = \lambda \in \mathbf{D}$.

By Theorem 2.3 in Ma (2012), there is a unique non-negative weak solution to (1.3) and the solution is a CBI-process with transition semigroup $(Q_t)_{t \geq 0}$ defined by (2.2). He also showed that there is a unique non-negative strong solution to (1.3).

With the conclusions above, The following theorem gives a necessary and sufficient condition for the ergodicity of the semigroup $(Q_t)_{t \geq 0}$.

Let $\kappa = \frac{b_{12}b_{21}}{b_{11}b_{22}}$. Then we get the following conclusion, the proof will be given in Section 5.

Theorem 2.2 Suppose both of the eigenvalues of $B = \begin{pmatrix} b_{11} & -b_{12} \\ -b_{21} & b_{22} \end{pmatrix}$ are positive, i.e. $\kappa < 1$. Then

the transition semigroup $(Q_t)_{t \geq 0}$ has the unique stationary distribution Q_∞ given by

$$\int_{\mathbf{D}} e^{-\langle y, \lambda \rangle} Q_\infty(dy) = \exp \left\{ - \int_0^\infty \langle A, v_s(\lambda) \rangle ds \right\}. \quad (2.4)$$

Moreover, for every $x \in \mathbf{D}$, we have $Q_t(x, \cdot) \rightarrow Q_\infty$ by weak convergence as $t \rightarrow \infty$.

With Theorem 2.2 the following corollary can be easily proved like the proof of Theorem 2.7 in Li and Ma (2013).

Corollary 2.3 *Suppose the conditions of Theorem 2.2 are satisfied, then X_t is mixing and its tail σ -algebra is trivial, e.g., Durrett (2010).*

The results of this paper on the asymptotics of the estimators will be derived under the assumption that the eigenvalues of B are all positive and X_0 is distributed according to the stationary law. By Birkhoff's ergodic theorem, X_t is a stationary and ergodic process, but by a fairly simple (continuous time) coupling argument it can be seen that they are valid for arbitrary initial distributions.

3 CLSEs and WCLSEs of (A, B) and Σ

Recall $\Delta = 1$. Let $\{\mathcal{F}_k = \sigma(\{X_0, \dots, X_k\}) : k = 0, 1, 2, \dots\}$. In this section we give the CLSEs and WCLSEs of the drift coefficients (A, B) and the diffusion coefficients Σ based on the observations $\{X_k : k = 0, \dots, n\}$.

Following Klimko and Nelson (1978) and Overbeck and Rydén (1997), we first define the CLSE of (A, B) . The basic ideas are explained as follows. By applying Itô's formula to (1.3), for any $t \geq r \geq 0$ we have

$$X_t = e^{-B(t-r)} X_r + \int_r^t e^{-B(t-s)} A ds + \int_r^t e^{-B(t-s)} \Sigma \sqrt{X_s} dW_s. \quad (3.1)$$

Apparently, $\int_r^t e^{-B(t-s)} \Sigma \sqrt{X_s} dW_s$ is a local martingale with respect to $\{\mathcal{F}_t\}$. For

$$\mathbf{E} \left[\int_0^t e^{-B(t-s)} \begin{pmatrix} \sigma_1^2 X_1(t) & 0 \\ 0 & \sigma_2^2 X_2(t) \end{pmatrix} e^{-B^T(t-s)} dt \right] < \infty, \quad \forall t > 0,$$

then $\int_r^t e^{-B(t-s)} \Sigma \sqrt{X_s} dW_s$ is a martingale.

Let \mathbf{I} be the identical matrix and define

$$\rho = B^{-1}(\mathbf{I} - e^{-B})A, \quad \gamma = e^{-B}. \quad (3.2)$$

From (3.1) and (3.2) we can easily obtain the stochastic regressive equation

$$X_k = \rho + \gamma X_{k-1} + \varepsilon_k, \quad (3.3)$$

where

$$\varepsilon_k = \int_{k-1}^k e^{-B(k-s)} \Sigma \sqrt{X_s} dW_s.$$

Apparently, $\{\varepsilon_k : k \geq 0\}$ is a martingale differential sequence with respect to $\{\mathcal{F}_k : k \geq 0\}$. Let $m(x; A, B) = \rho + \gamma x$, then

$$\mathbf{E}_\theta[X_k | X_{k-1}] = m(X_{k-1}; A, B).$$

Suppose $g(x)$ is a Borel measurable function on \mathbb{R}^2 , satisfying that

$$\mathbf{E}(|g(X_0)|^2) < \infty,$$

let $g_k = g(X_{k-1})$, where $k = 1, \dots, n$. The WCLSE of (A, B) can be given by minimizing the sum of squares

$$\sum_{k=1}^n g_k \varepsilon_k^T \varepsilon_k = \sum_{k=1}^n g_k (X_k - \rho - \gamma X_{k-1})^T (X_k - \rho - \gamma X_{k-1}). \quad (3.4)$$

In particular, the estimators of (ρ, γ) are given by

$$\hat{\rho}_n = \bar{g}_n (\bar{X}_n - \hat{\gamma}_n \bar{X}_n) \quad (3.5)$$

and

$$\hat{\gamma}_n = (\bar{T}_{1,n} - \tilde{T}_{1,n}) \cdot (\bar{T}_{2,n} - \tilde{T}_{2,n})^{-1}, \quad (3.6)$$

where

$$\begin{aligned} \bar{X}_n &= \frac{1}{n} \sum_{k=1}^n g_k X_k, \quad \tilde{X}_n = \frac{1}{n} \sum_{k=1}^n g_k X_{k-1}, \quad \bar{g}_n = \frac{1}{n} \sum_{k=1}^n g_k, \\ \bar{T}_{1,n} &= \frac{1}{n} \sum_{k=1}^n (g_k X_k - \bar{X}_n)(g_k X_{k-1} - \tilde{X}_n)^T, \\ \tilde{T}_{1,n} &= \frac{1}{n} \sum_{k=1}^n (g_k - \bar{g}_n)(g_k X_k X_{k-1}^T - \frac{1}{n} \sum_{k=1}^n g_k X_k X_{k-1}^T), \\ \bar{T}_{2,n} &= \frac{1}{n} \sum_{k=1}^n (g_k X_{k-1} - \tilde{X}_n)(g_k X_{k-1} - \tilde{X}_n)^T, \\ \tilde{T}_{2,n} &= \frac{1}{n} \sum_{k=1}^n (g_k - \bar{g}_n)(g_k X_{k-1} X_{k-1}^T - \frac{1}{n} \sum_{k=1}^n g_k X_{k-1} X_{k-1}^T). \end{aligned}$$

According to (3.2), we can easily get the CLSE of (A, B) by simple calculation,

$$\begin{aligned} \hat{A}_n &= (\mathbf{I} - \hat{\gamma}_n)^{-1} \hat{B}_n \hat{\rho}_n, \\ \hat{B}_n &= -\ln \hat{\gamma}_n, \end{aligned}$$

Where $\ln \hat{\gamma}_n = \sum_{k=0}^{\infty} (-1)^{k-1} \frac{1}{k} (\hat{\gamma}_n - 1)^k$. From the proof of Theorem 4.1, $\hat{\gamma}_n \xrightarrow{\text{a.s.}} \gamma$ and the spectral radius of γ is less than one, so \hat{B}_n is reasonable.

We now turn to the estimation of Σ . For the connection between Σ and (σ_1, σ_2) , we just need to study (σ_1, σ_2) . For this, we need to compute

$$v(X_{k-1}; \theta) = \mathbf{E}_\theta[(X_k - m(X_{k-1}; A, B))(X_k - m(X_{k-1}; A, B))^T | X_{k-1}]. \quad (3.7)$$

By applying Itô's formula to $X_t X_t^T$, for any $t \geq r \geq 0$ we have

$$\begin{aligned} X_t X_t^T &= X_r X_r^T + \int_r^t [X_s A^T + A X_s^T + \Sigma \sqrt{X_s} \sqrt{X_s}^T \Sigma^T - (X_s X_s^T B^T - B X_s X_s^T)] ds \\ &\quad + \int_r^t \Sigma \sqrt{X_s} dW_s X_s^T + \int_r^t X_s dW_s^T \sqrt{X_s}^T \Sigma^T, \end{aligned}$$

where

$$\Sigma \sqrt{X_s} \sqrt{X_s}^T \Sigma^T = \begin{pmatrix} \sigma_1^2 X_1(s) & 0 \\ 0 & \sigma_2^2 X_2(s) \end{pmatrix}.$$

Let $f(s) = \mathbf{E}_\theta[X_{k-1+s} | X_{k-1}]$ and $h(s) = \mathbf{E}_\theta[X_{k-1+s} X_{k-1+s}^T | X_{k-1}]$, $s \geq 0$. Then for any $t \geq 0$

$$h(t) = X_{k-1} X_{k-1}^T + \int_0^t \left\{ f(s) A^T + A f(s)^T - (h(s) B^T - B h(s)) + \begin{pmatrix} \sigma_1^2 f_1(s) & 0 \\ 0 & \sigma_2^2 f_2(s) \end{pmatrix} \right\} ds. \quad (3.8)$$

We can get $h(t)$ by solving (3.8),

$$\begin{aligned} h(t) &= e^{-Bt} X_{k-1} X_{k-1}^T e^{-B^T t} + \int_0^t e^{-B(t-s)} (f(s) A^T + A f(s)^T) e^{-B^T(t-s)} ds \\ &\quad + \int_0^t e^{-B(t-s)} \begin{pmatrix} \sigma_1^2 f_1(s) & 0 \\ 0 & \sigma_2^2 f_2(s) \end{pmatrix} e^{-B^T(t-s)} ds. \end{aligned}$$

For

$$f(t) = e^{-Bt} X_{k-1} + \int_0^t e^{-B(t-s)} A ds,$$

we can easily get

$$f(t) f(t)^T = e^{-Bt} X_{k-1} X_{k-1}^T e^{-B^T t} + \int_0^t e^{-B(t-s)} (f(s) A^T + A f(s)^T) e^{-B^T(t-s)} ds.$$

Thus

$$v(X_{k-1}; \theta) = h(1) - f(1) f(1)^T = \sigma_1^2 \eta_1 + \sigma_2^2 \eta_2, \quad (3.9)$$

where

$$\begin{aligned} \eta_1 &= \int_0^1 e^{-B(1-s)} \begin{pmatrix} f_1(s) & 0 \\ 0 & 0 \end{pmatrix} e^{-B^T(1-s)} ds, \\ \eta_2 &= \int_0^1 e^{-B(1-s)} \begin{pmatrix} 0 & 0 \\ 0 & f_2(s) \end{pmatrix} e^{-B^T(1-s)} ds. \end{aligned}$$

Before giving the estimator, we give the following important definition.

Definition 3.1 Assume $A = (a_1, \dots, a_m)$ is $n \times m$ matrix, where $\{a_i : i = 1, \dots, m\}$ are $n \times 1$ vectors, and define a operator Vec from a space of matrixes to a space of vectors, i.e.

$$\text{Vec}(A) = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}.$$

Now we consider (3.9) as a regression equation with $(X_k - m(X_{k-1}; A, B))(X_k - m(X_{k-1}; A, B))^T$ being the response, $\sigma_1^2 \eta_1 + \sigma_2^2 \eta_2$ being the predictor, and (σ_1^2, σ_2^2) being the unknown. Recall $\{g_k : k = 1, \dots, n\}$. The WCLSE of (σ_1^2, σ_2^2) can be the minimizer of the following sum of squares, here we must of course substitute (A, B) by the corresponding estimates

$$\sum_{k=1}^n g_k \text{Vec}(Z_k - (\sigma_1^2 \hat{\eta}_{1,n} + \sigma_2^2 \hat{\eta}_{2,n}))^T \text{Vec}(Z_k - (\sigma_1^2 \hat{\eta}_{1,n} + \sigma_2^2 \hat{\eta}_{2,n})),$$

where $(\hat{\eta}_{1,n}, \hat{\eta}_{2,n})$ is (η_1, η_2) with (A, B) substituted by (\hat{A}_n, \hat{B}_n) and

$$Z_k = (X_k - m(X_{k-1}; \hat{A}_n, \hat{B}_n))(X_k - m(X_{k-1}; \hat{A}_n, \hat{B}_n))^T.$$

In particular, those estimators of σ_1^2 and σ_2^2 are given by

$$\hat{\sigma}_{1,n}^2 = \frac{\varphi_{11,n} - \varphi_{12,n}}{\psi_n} \quad (3.10)$$

and

$$\hat{\sigma}_{2,n}^2 = \frac{\varphi_{21,n} - \varphi_{22,n}}{\psi_n}, \quad (3.11)$$

where $(\hat{\eta}_{1,n}, \hat{\eta}_{2,n})$ is (η_1, η_2) with (A, B) substituted by (\hat{A}_n, \hat{B}_n)

$$\varphi_{11,n} = \frac{1}{n} \sum_{k=1}^n g_k \text{Vec}(Z_k)^T \text{Vec}(\hat{\eta}_{1,n}) \text{Vec}(\hat{\eta}_{2,n})^T \text{Vec}(\hat{\eta}_{2,n}), \quad (3.12)$$

$$\varphi_{12,n} = \frac{1}{n} \sum_{k=1}^n g_k \text{Vec}(Z_k)^T \text{Vec}(\hat{\eta}_{2,n}) \text{Vec}(\hat{\eta}_{1,n})^T \text{Vec}(\hat{\eta}_{2,n}), \quad (3.13)$$

$$\varphi_{21,n} = \frac{1}{n} \sum_{k=1}^n g_k \text{Vec}(Z_k)^T \text{Vec}(\hat{\eta}_{2,n}) \text{Vec}(\hat{\eta}_{1,n})^T \text{Vec}(\hat{\eta}_{1,n}), \quad (3.14)$$

$$\varphi_{22,n} = \frac{1}{n} \sum_{k=1}^n g_k \text{Vec}(Z_k)^T \text{Vec}(\hat{\eta}_{1,n}) \text{Vec}(\hat{\eta}_{1,n})^T \text{Vec}(\hat{\eta}_{2,n}) \quad (3.15)$$

and

$$\psi_n = \bar{g}_n \left[\text{Vec}(\hat{\eta}_{1,n})^T \text{Vec}(\hat{\eta}_{1,n}) \text{Vec}(\hat{\eta}_{2,n})^T \text{Vec}(\hat{\eta}_{2,n}) - \left(\text{Vec}(\hat{\eta}_{1,n})^T \text{Vec}(\hat{\eta}_{2,n}) \right)^2 \right]. \quad (3.16)$$

If choose $g(x) \equiv 1$, we will get the CLSEs respectively. So in the later of this paper, we just need to deal with WCLSEs. If the observations of X_t are too large, we can choose $g(x) = \frac{1}{1+|x|}$, where $|x|$ is the norm of x , this WCLSEs are usually discussed in other papers for its good properties.

4 Consistency and asymptotic normality of (\hat{A}_n, \hat{B}_n) and $(\hat{\sigma}_{1,n}^2, \hat{\sigma}_{2,n}^2)$

In this section we devote to show that (\hat{A}_n, \hat{B}_n) and $(\hat{\sigma}_{1,n}^2, \hat{\sigma}_{2,n}^2)$ are strongly consistent. Further, we also analyze the central limit theorem of the WCLSEs. All the proofs will be given in Section 5.

Theorem 4.1 *The estimator (\hat{A}_n, \hat{B}_n) is strongly consistent, i.e. $(\hat{A}_n, \hat{B}_n) \xrightarrow{\text{a.s.}} (A, B)$ as $n \rightarrow \infty$.*

Remark 4.2 *The estimators \hat{A}_n , and \hat{B}_n derived solely from the conditional mean function are robust against misspecification of the diffusion term in (1.3).*

Remark 4.3 *In the estimation of (A, B) , the diffusion term in (1.3) can be replaced by $\sigma(X_t)dW_t$, where $\sigma(\cdot)$ is an arbitrary function such that the induced stationary distribution of $\{X_t\}$ has finite second moment.*

Theorem 4.4 *The estimator $(\hat{\sigma}_{1,n}^2, \hat{\sigma}_{2,n}^2)$ are strongly consistent, i.e. $(\hat{\sigma}_{1,n}^2, \hat{\sigma}_{2,n}^2) \xrightarrow{\text{a.s.}} (\sigma_1^2, \sigma_2^2)$ as $n \rightarrow \infty$.*

Remark 4.5 *From the proof of Theorem 4.4 in Section 5, we can see any weakly consistent estimator of (A, B) gives weakly consistent estimator of Σ .*

Let $\hat{\theta}_n = (\hat{a}_{1,n}, \hat{a}_{2,n}, \hat{b}_{11,n}, \hat{b}_{12,n}, \hat{b}_{21,n}, \hat{b}_{22,n}, \hat{\sigma}_{1,n}, \hat{\sigma}_{2,n})$ is the WCLSE of θ given in the last section. Now we analyze its asymptotic normality.

According to the argument above, we know $\hat{\theta}_n$ is the unique solution of the following equation:

$$G_n(\theta) = \sum_{k=1}^n g_k \{ w_0(X_{k-1}; \theta) (X_k - \rho - \gamma X_{k-1}) + w_1(X_{k-1}; \theta) \text{Vec}(Z_k - (\sigma_1^2 \eta_1 + \sigma_2^2 \eta_2)) \} = 0, \quad (4.1)$$

where

$$w_0(x; \theta) = \begin{pmatrix} \frac{\partial \rho_1}{\partial a_1} & \frac{\partial \rho_2}{\partial a_1} \\ \frac{\partial \rho_1}{\partial a_2} & \frac{\partial \rho_2}{\partial a_2} \\ \frac{\partial \rho_1}{\partial b_{11}} + \left(\frac{\partial \gamma_{11}}{\partial b_{11}}, \frac{\partial \gamma_{12}}{\partial b_{11}} \right) x & \frac{\partial \rho_2}{\partial b_{11}} + \left(\frac{\partial \gamma_{21}}{\partial b_{11}}, \frac{\partial \gamma_{22}}{\partial b_{11}} \right) x \\ \frac{\partial \rho_1}{\partial b_{12}} + \left(\frac{\partial \gamma_{11}}{\partial b_{12}}, \frac{\partial \gamma_{12}}{\partial b_{12}} \right) x & \frac{\partial \rho_2}{\partial b_{12}} + \left(\frac{\partial \gamma_{21}}{\partial b_{12}}, \frac{\partial \gamma_{22}}{\partial b_{12}} \right) x \\ \frac{\partial \rho_1}{\partial b_{21}} + \left(\frac{\partial \gamma_{11}}{\partial b_{21}}, \frac{\partial \gamma_{12}}{\partial b_{21}} \right) x & \frac{\partial \rho_2}{\partial b_{21}} + \left(\frac{\partial \gamma_{21}}{\partial b_{21}}, \frac{\partial \gamma_{22}}{\partial b_{21}} \right) x \\ \frac{\partial \rho_1}{\partial b_{22}} + \left(\frac{\partial \gamma_{11}}{\partial b_{22}}, \frac{\partial \gamma_{12}}{\partial b_{22}} \right) x & \frac{\partial \rho_2}{\partial b_{22}} + \left(\frac{\partial \gamma_{21}}{\partial b_{22}}, \frac{\partial \gamma_{22}}{\partial b_{22}} \right) x \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

and

$$w_1(x; \theta) = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \text{Vec}(\eta_1) & \text{Vec}(\eta_2) \end{pmatrix}^T.$$

The key to the analysis of $\hat{\theta}_n$ is that $G_n(\theta)$ is a \mathbb{P}_θ -martingale with respect to $\{\mathcal{F}_n : n \geq 0\}$. It's obvious, so we will not give the proof.

Let

$$\mu_3(x; \theta) = \mathbf{E}_\theta[(X_k - m(X_{k-1}; \theta))\text{Vec}(Z_k - v(x; \theta))^T | X_{k-1} = x]$$

and

$$\mu_4(x; \theta) = \mathbf{E}_\theta[\text{Vec}(Z_k - v(x; \theta))\text{Vec}(Z_k - v(x; \theta))^T | X_{k-1} = x].$$

Theorem 4.6 *We have $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V^{-1}WV^{-T})$ as $n \rightarrow \infty$, where*

$$\begin{aligned} W = \mathbf{E}_\theta \Big[& g_1^2 \{ w_0(X_0; \theta) v(X_0; \theta) w_0^T(X_0; \theta) + w_1(X_0; \theta) \mu_4(X_0; \theta) w_1^T(X_0; \theta) \\ & + w_0(X_0; \theta) \mu_3(X_0; \theta) w_1^T(X_0; \theta) + w_1(X_0; \theta) \mu_3^T(X_0; \theta) w_0^T(X_0; \theta) \} \Big] \end{aligned} \quad (4.2)$$

and

$$V = -\mathbf{E}_\theta \left[g_1 \left\{ w_0(X_0; \theta) \left(\frac{\partial m(X_0; \theta)}{\partial \theta} \right) + w_1(X_0; \theta) \left(\frac{\partial}{\partial \theta} \text{Vec}(v(X_0; \theta)) \right) \right\} \right]. \quad (4.3)$$

By taking $g \equiv 1$ one can see that all the conclusions above also hold for the CLSEs.

5 Proofs

In this section, we will give the proofs for the theorems in Section 2 and 4

Proof of Theorem 2.2 Let $u(t, \lambda)$ be the unique solution of the linear equation:

$$\begin{cases} \frac{\partial}{\partial t} u_1(t, \lambda) = -b_{11}u_1(t, \lambda) + b_{12}u_2(t, \lambda), \\ \frac{\partial}{\partial t} u_2(t, \lambda) = b_{21}u_1(t, \lambda) - b_{22}u_2(t, \lambda), \end{cases} \quad (5.1)$$

with the initial condition $u(0, \lambda) = \lambda \in \mathbf{D}$. By applying the comparison theorem to (2.3) and (5.1), we have $v_i(t, \lambda) \leq u_i(t, \lambda)$ for $t \geq 0$, where $i = 1, 2$. By solving (5.1), we get $u(t, \lambda) = e^{-Bt}\lambda$. Let $\xi_1, \xi_2 > 0$ be the eigenvalues of B . If $\xi_1 = \xi_2$, then $b_{11} = b_{22} > 0$, $b_{12} = b_{21} = 0$ and

$$u(t, \lambda) = \begin{pmatrix} e^{-b_{11}t} \lambda_1 \\ e^{-b_{22}t} \lambda_2 \end{pmatrix}. \quad (5.2)$$

If $\xi_1 \neq \xi_2$, let $L_i = \{\ell : (B - \xi_i \mathbf{I})\ell = 0\}$, where $i = 1, 2$. Then V_1, V_2 are subspaces of \mathbb{R}^2 . For any $\lambda \in \mathbf{D}$, there exist $\ell_1 \in L_1$ and $\ell_2 \in L_2$, such that $\lambda = \ell_1 + \ell_2$ and

$$(B - \xi_i \mathbf{I})^l \ell_i = 0 \quad l \geq 1, \quad i = 1, 2. \quad (5.3)$$

By (5.3), we have

$$e^{-Bt}\lambda = e^{-\xi_1 t} \ell_1 + e^{-\xi_2 t} \ell_2. \quad (5.4)$$

By (5.2) and (5.4), we have

$$v_i(t, \lambda) \leq u_i(t, \lambda) \leq C_1(\lambda) e^{-(\xi_1 \wedge \xi_2)t} \quad \text{for some } C_1(\lambda) > 0. \quad (5.5)$$

Thus $\lim_{t \rightarrow +\infty} v_i(t, \lambda) = 0$. So for $\forall x \in \mathbf{D}$,

$$\lim_{t \rightarrow +\infty} \int_{\mathbf{D}} e^{-\langle y, \lambda \rangle} Q_t(x, dy) = \lim_{t \rightarrow +\infty} e^{-\langle x, v_t(\lambda) \rangle - \int_0^t \langle A, v_s(\lambda) \rangle ds} = e^{-\int_0^\infty \langle A, v_s(\lambda) \rangle ds}.$$

By (5.5), we have $\langle A, v_s(\lambda) \rangle = a_1 v_1(s, \lambda) + a_2 v_2(s, \lambda) \leq C_2(\lambda) e^{-(\xi_1 \wedge \xi_2)t}$, for some $C_2(\lambda) > 0$. Then

$$\int_0^\infty \langle A, v_s(\lambda) \rangle ds < \infty \quad \text{for } \forall \lambda \in \mathbf{D}.$$

Thus, there is a unique distribution $Q_\infty(\cdot)$ satisfying

$$\int_{\mathbf{D}} e^{-\langle y, \lambda \rangle} Q_\infty(dy) = \exp \left\{ - \int_0^\infty \langle A, v_s(\lambda) \rangle ds \right\}.$$

□

Proof of Theorem 4.1 By the relationship of (\hat{A}_n, \hat{B}_n) and $(\hat{\rho}_n, \hat{\gamma}_n)$, we just need to prove the consistency of $(\hat{\rho}_n, \hat{\gamma}_n)$, i.e.

$$(\hat{\rho}_n, \hat{\gamma}_n) \xrightarrow{\text{a.s.}} (\rho, \gamma) \text{ as } n \rightarrow \infty.$$

By ergodicity, we have

$$\bar{T}_{1,n} \xrightarrow{\text{a.s.}} \text{Cov}(g_1 X_1, g_1 X_0) \text{ as } n \rightarrow \infty$$

and

$$\tilde{T}_{1,n} \xrightarrow{\text{a.s.}} \text{Cov}(g_1, g_1 X_1 X_0^T) \text{ as } n \rightarrow \infty.$$

For

$$\begin{aligned} \text{Cov}(g_1 X_1, g_1 X_0) &= \mathbf{E}_\theta[g_1^2 \mathbf{E}_\theta(X_1 | \mathcal{F}_0) X_0^T] - \mathbf{E}_\theta[g_1 \mathbf{E}_\theta(X_1 | \mathcal{F}_0)] \mathbf{E}_\theta[g_1 X_0^T], \\ \text{Cov}(g_1, g_1 X_1 X_0^T) &= \mathbf{E}_\theta[g_1^2 \mathbf{E}_\theta(X_1 | \mathcal{F}_0) X_0^T] - \mathbf{E}_\theta[g_1] \mathbf{E}_\theta[g_1 \mathbf{E}_\theta(X_1 | \mathcal{F}_0) X_0^T], \end{aligned}$$

then we get

$$\bar{T}_{1,n} - \tilde{T}_{1,n} \xrightarrow{\text{a.s.}} \gamma \{ \mathbf{E}_\theta[g_1] \mathbf{E}_\theta[g_1 X_0 X_0^T] - \mathbf{E}_\theta[g_1 X_0] \mathbf{E}_\theta[g_1 X_0^T] \} \text{ as } n \rightarrow \infty. \quad (5.6)$$

Similarly, we can easily get

$$\bar{T}_{2,n} - \tilde{T}_{2,n} \xrightarrow{\text{a.s.}} \mathbf{E}_\theta[g_1] \mathbf{E}_\theta[g_1 X_0 X_0^T] - \mathbf{E}_\theta[g_1 X_0] \mathbf{E}_\theta[g_1 X_0^T] \text{ as } n \rightarrow \infty. \quad (5.7)$$

By (3.6), (5.6) and (5.7),

$$\hat{\gamma}_n \xrightarrow{\text{a.s.}} \gamma \text{ as } n \rightarrow \infty.$$

Similarly, by ergodicity and (3.5),

$$\hat{\rho}_n \xrightarrow{\text{a.s.}} \rho \text{ as } n \rightarrow \infty.$$

□

Proof of Theorem 4.4 By ergodicity and theorem 4.1, we can get

$$(\text{Vec}(\hat{\eta}_{1,n}), \text{Vec}(\hat{\eta}_{2,n})) \xrightarrow{\text{a.s.}} (\text{Vec}(\eta_1), \text{Vec}(\eta_2)) \text{ as } n \rightarrow \infty.$$

Next we will prove ψ_n and $\varphi_{ij,n}$ converge almost surly, where $i, j = 1, 2$. Fix θ . For $\forall \vartheta \in [0, \infty)^8$, let

$$h(x, y; \vartheta) = g(x) \text{Vec}((y - m(x; \vartheta))(y - m(x; \vartheta))^T)^T \text{Vec}(\eta_1(\vartheta)) \text{Vec}(\eta_2(\vartheta))^T \text{Vec}(\eta_2(\vartheta))$$

and $U \subset [0, +\infty)^8$ be a neighborhood of θ such that

$$\mathbf{E}_\theta[\sup_{\vartheta \in U} |h(X_0, X_1; \vartheta)|] < \infty.$$

By ergodicity, (3.12) and Theorem 4.1

$$\overline{\lim}_{n \rightarrow \infty} \varphi_{11,n} \leq \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \sup_{\vartheta \in U} h(X_{k-1}, X_k; \vartheta) = \mathbf{E}_\theta[\sup_{\vartheta \in U} h(X_0, X_1; \vartheta)].$$

Let $U \downarrow \{\theta\}$. Then

$$\overline{\lim}_{n \rightarrow \infty} \varphi_{11,n} \leq \mathbf{E}_\theta[h(X_0, X_1; \theta)]. \quad (5.8)$$

Similarly, we can prove

$$\underline{\lim}_{n \rightarrow \infty} \varphi_{11,n} \geq \mathbf{E}_\theta[h(X_0, X_1; \theta)]. \quad (5.9)$$

By (5.8) and (5.9),

$$\begin{aligned} \varphi_{11,n} &\xrightarrow{\text{a.s.}} \sigma_1^2 \mathbf{E}_\theta [g_1 \text{Vec}(\eta_1)^T \text{Vec}(\eta_1) \text{Vec}(\eta_2)^T \text{Vec}(\eta_2)] \\ &\quad + \sigma_2^2 \mathbf{E}_\theta [g_1 \text{Vec}(\eta_2)^T \text{Vec}(\eta_1) \text{Vec}(\eta_2)^T \text{Vec}(\eta_2)] \text{ as } n \rightarrow \infty. \end{aligned} \quad (5.10)$$

Similarly, when $n \rightarrow \infty$ we can get

$$\varphi_{12,n} \xrightarrow{\text{a.s.}} \sigma_1^2 \mathbf{E}_\theta [g_1 (\text{Vec}(\eta_1)^T \text{Vec}(\eta_2))^2] + \sigma_2^2 \mathbf{E}_\theta [g_1 \text{Vec}(\eta_2)^T \text{Vec}(\eta_1) \text{Vec}(\eta_2)^T \text{Vec}(\eta_2)] \quad (5.11)$$

and

$$\psi_n \xrightarrow{\text{a.s.}} \mathbf{E}_\theta [g_1 \text{Vec}(\eta_1)^T \text{Vec}(\eta_1) \text{Vec}(\eta_2)^T \text{Vec}(\eta_2) - g_1 (\text{Vec}(\eta_1)^T \text{Vec}(\eta_2))^2]. \quad (5.12)$$

By (3.10), (5.10), (5.11) and (5.12),

$$\hat{\sigma}_{1,n}^2 \xrightarrow{\text{a.s.}} \sigma_1^2 \text{ as } n \rightarrow \infty.$$

Similarly, we can prove

$$\hat{\sigma}_{2,n}^2 \xrightarrow{\text{a.s.}} \sigma_2^2 \text{ as } n \rightarrow \infty.$$

□

Proof of Theorem 4.6 Fix θ . By Theorem 4.1 and 4.4, making a Taylor expansion of G_n about θ , it is enough to show that

$$G_n(\theta') = G_n(\theta) + G_n'(\theta)(\theta' - \theta) + \frac{1}{2} G_n''(\xi)(\theta' - \theta)(\theta' - \theta)^T, \quad (5.13)$$

where ξ is between θ' and θ . Each element of (5.13) is

$$G_{n,i}(\theta') = G_{n,i}(\theta) + \sum_{l=1}^8 \frac{\partial}{\partial \theta_l} G_{n,i}(\theta)(\theta'_l - \theta_l) + \frac{1}{2} \sum_{l=1}^8 \sum_{j=1}^8 \frac{\partial^2}{\partial \theta_l \partial \theta_j} G_{n,i}(\xi)(\theta'_j - \theta_j)(\theta'_l - \theta_l),$$

where $i = 1, \dots, 8$.

If choose $\theta' = \hat{\theta}_n$ in the above equation, we will get

$$-\frac{1}{\sqrt{n}} G_{n,i}(\theta) = \sum_{l=1}^8 \sqrt{n}(\hat{\theta}_{n,l} - \theta_l) \left\{ \frac{1}{n} \frac{\partial}{\partial \theta_l} G_{n,i}(\theta) + \frac{1}{2n} \sum_{j=1}^8 \frac{\partial^2}{\partial \theta_l \partial \theta_j} G_{n,i}(\xi)(\hat{\theta}_{n,j} - \theta_j) \right\}. \quad (5.14)$$

Let

$$\begin{aligned} T_{n,l} &= \sqrt{n}(\hat{\theta}_{n,l} - \theta_l), \quad Y_{n,i} = -\frac{1}{\sqrt{n}} G_{n,i}(\theta), \\ D_{n,il} &= \frac{1}{n} \frac{\partial}{\partial \theta_l} G_{n,i}(\theta) + \frac{1}{2n} \sum_{j=1}^8 \frac{\partial^2}{\partial \theta_l \partial \theta_j} G_{n,i}(\xi)(\hat{\theta}_{n,j} - \theta_j). \end{aligned}$$

Then (5.14) turns to be

$$D_n T_n = Y_n, \quad (5.15)$$

where $D_n = (D_{n,il})_{8 \times 8}$, $T_n = (T_{n,1}, \dots, T_{n,8})^T$, $Y_n = (Y_{n,1}, \dots, Y_{n,8})^T$. Next we will prove the following results respectively,

$$D_n \xrightarrow{\text{a.s.}} V \text{ as } n \rightarrow \infty \quad (5.16)$$

and

$$Y_n \xrightarrow{d} \mathcal{N}(0, W) \text{ as } n \rightarrow \infty. \quad (5.17)$$

By (5.15), (5.16) and (5.17), we can get

$$T_n \xrightarrow{d} \mathcal{N}(0, V^{-1}WV^{-T}) \text{ as } n \rightarrow \infty.$$

(1) Firstly, we prove (5.16) holds. Let

$$h(x, y; \theta) = w_0(x; \theta)(y - m(x; \theta)) + w_1(x; \theta)\text{Vec}(z - v(x; \theta)),$$

where $z = (y - m(x; A, B))(y - m(x; A, B))^T$. Then (4.1) turns out to be

$$G_n(\theta) = \sum_{k=1}^n g_k h(X_{k-1}, X_k; \theta).$$

It's easy to prove that $\exists H(x, y)$ and $\theta \in U$, such that

- (i) $|\frac{\partial^2}{\partial \theta_j \partial \theta_l} h_i(x, y; \theta)| \leq H(x, y)$ uniformly in U , for $i, j, l = 1, \dots, 8$.
- (ii) $E_\theta[|H(X_0, X_1)|] < \infty$.

Then

$$\begin{aligned} \left| \frac{1}{2n} \sum_{j=1}^8 \frac{\partial^2}{\partial \theta_l \partial \theta_j} G_{n,i}(\xi)(\hat{\theta}_{n,j} - \theta_j) \right| &= \left| \frac{1}{2n} \sum_{j=1}^8 (\hat{\theta}_{n,j} - \theta_j) \sum_{k=1}^n g_k \frac{\partial^2}{\partial \theta_l \partial \theta_j} h_i(X_{k-1}; X_k; \xi) \right| \\ &\leq \sum_{j=1}^8 |\hat{\theta}_{n,j} - \theta_j| \frac{1}{2n} \sum_{k=1}^n |g_k| H(X_{k-1}, X_k). \end{aligned}$$

By Theorem 4.1, 4.4, and

$$\frac{1}{2n} \sum_{k=1}^n |g_k| H(X_{k-1}, X_k) \xrightarrow{\text{a.s.}} \frac{1}{2} E_\theta[|g_1| H(X_0, X_m)] < \infty \text{ as } n \rightarrow \infty,$$

we can get

$$\left| \frac{1}{2n} \sum_{j=1}^8 \frac{\partial^2}{\partial \theta_l \partial \theta_j} G_{n,i}(\xi)(\hat{\theta}_{n,j} - \theta_j) \right| \xrightarrow{\text{a.s.}} 0 \text{ as } n \rightarrow \infty.$$

By (4.1),

$$\begin{aligned} \frac{1}{n} \frac{\partial}{\partial \theta_l} G_{n,i}(\theta) &= \frac{1}{n} \sum_{k=1}^n g_k \left\{ \left(\frac{\partial}{\partial \theta_l} w_{0,i}(X_{k-1}; \theta) \right) (X_k - m(X_{k-1}; \theta)) \right. \\ &\quad + w_{0,i}(X_{k-1}; \theta) \left(-\frac{\partial}{\partial \theta_l} m(X_{k-1}; \theta) \right) \\ &\quad + \left(\frac{\partial}{\partial \theta_l} w_{1,i}(X_{k-1}; \theta) \right) \text{Vec}(Z_k - v(X_{k-1}; \theta))^T \\ &\quad + w_{1,i}(X_{k-1}; \theta) \text{Vec} \left(\left(-\frac{\partial}{\partial \theta_l} m(X_{k-1}; \theta) \right) (X_k - m(X_{k-1}; \theta))^T \right. \\ &\quad \left. \left. + (X_k - m(X_{k-1}; \theta)) \left(-\frac{\partial}{\partial \theta_l} m(X_{k-1}; \theta) \right)^T - \frac{\partial}{\partial \theta_l} v(X_{k-1}; \theta) \right) \right\}, \end{aligned}$$

where $\{w_{0,i}, w_{1,i} : i = 1, \dots, 8\}$ are the row vectors of w_0 and w_1 . By ergodicity, we calculate each part respectively, we can get

$$\frac{1}{n} \frac{\partial}{\partial \theta_l} G_{n,i}(\theta) \rightarrow V_{i,l},$$

where

$$V_{i,l} = -E_\theta \left[g_1 \left[w_{0,i}(X_0; \theta) \frac{\partial}{\partial \theta_l} m(X_0; \theta) + w_{1,i}(X_0; \theta) \text{Vec} \left(\frac{\partial}{\partial \theta_l} v(X_0; \theta) \right) \right] \right].$$

Thus we get (5.16).

(2) Secondly, we prove (5.17) holds. Recall $\{G_n(\theta)\}$ is a martingale, let

$$\begin{aligned} \frac{1}{n} V_n &= \frac{1}{n} \sum_{k=1}^n E_\theta [(g_k h(X_{k-1}, X_k; \theta))(g_k h(X_{k-1}, X_k; \theta))^T | \mathcal{F}_{k-1}] \\ &= \frac{1}{n} \sum_{k=1}^n g_k^2 [w_0(X_{k-1}; \theta) v(X_{k-1}; \theta) w_0^T(X_{k-1}; \theta) + w_1(X_{k-1}; \theta) \mu_4(X_{k-1}; \theta) w_1^T(X_{k-1}; \theta) \\ &\quad + w_0(X_{k-1}; \theta) \mu_3(X_{k-1}; \theta) w_1^T(X_{k-1}; \theta) + w_1(X_{k-1}; \theta) \mu_3^T(X_{k-1}; \theta) w_0^T(X_{k-1}; \theta)]. \end{aligned}$$

By ergodicity, we can easily get

$$\frac{1}{n} V_n \rightarrow W \text{ as } n \rightarrow \infty.$$

By stationarity, For $\forall \varepsilon > 0$

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n E_\theta [g_k^2 h(X_{k-1}, X_k; \theta)^T h(X_{k-1}, X_k; \theta) 1_{\{g_k^2 h(X_{k-1}, X_k; \theta)^T h(X_{k-1}, X_k; \theta) > n\varepsilon^2\}}] \\ = E_\theta [(g_1^2 h(X_0, X_1; \theta)^T h(X_0, X_1; \theta) 1_{\{g_1^2 h(X_0, X_1; \theta)^T h(X_0, X_1; \theta) > n\varepsilon^2\}})] \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Thus, by the martingale central theorem (see, e.g., Durrett, 2010), (5.17) holds. □

References

- [1] Barczy, M.; Döring, L.; Li, Z. and Pap, G. (2013a): On parameter estimation for critical affine processes. *Electron. J. Statist.* **7** 647–696.
- [2] Barczy, M.; Döring, L.; Li, Z. and Pap, G. (2013b): Ergodicity for an affine two factor model. Preprint (arXiv:1302.2534).
- [3] Barczy, M.; Döring, L.; Li, Z. and Pap, G. (2013c): Parameter estimation for an affine two factor model. Preprint (arXiv:1302.3451).
- [4] Bibby, B.M. and Sørensen, M. (1995): Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1**, 17–39.
- [5] Cox, J.; Ingersoll, J. and Ross, S. (1985): A theory of the term structure of interest rate. *Econometrica* **53**, 385–408.
- [6] Dawson, D.A. and Li, Z. (2006): Skew convolution semigroups and affine Markov processes. *Ann. Probab.* **34**, 1103–1142.
- [7] Duffie, D.; Filipović, D. and Schachermayer, W. (2003): Affine processes and applications in finance. *Annal. Appl. Probab.* **13**, 984–1053.
- [8] Durrett, R. (2010): *Probability: Theory and Examples*. Cambridge University Press.
- [9] Ikeda, N. and Watanabe, S. (1989): *Stochastic Differential Equations and Diffusion Processes*. North-Holland Kodansha, Amsterdam/Tokyo.
- [10] Jiřina, M. (1958): Stochastic branching processes with continuous state space. *Czech. Math. J.* **8**, 292–313.
- [11] Kawazu, K. and Watanabe, S. (1971): Branching processes with immigration and related limit theorems. *Theory Probab. Appl.* **16**, 36–54.
- [12] Klimko, L.A. and Nelson, P.I. (1978): On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629–642.

- [13] Li, Z. (2011): *Measure-Valued Branching Markov Processes*. Springer, Berlin.
- [14] Li, Z. and Ma, C. (2013): Asymptotic properties of estimators in a stable Cox-Ingersoll-Ross model. Preprint (arXiv:1301.3243).
- [15] Longstaff, F.A. and Schwartz, E.S. (1992): Interest rate volatility and the term structure: A two-factor general equilibrium model. *J. Finance* **47**, 1259-1282.
- [16] Ma, R. (2012): Stochastic equations for two-type continuous-state branching processes with immigration. *Acta Math. Sinica* **29**, 287-294.
- [17] Overbeck, L. and Rydén, T. (1997): Estimation in the Cox-Ingersoll-Ross model. *Econometric Theory* **13**, 430-461.
- [18] Quine, M.P. and Durham, P. (1977): Estimation for multitype branching processes. *J. Appl. Probab.* **14**, 829-835.
- [19] Revuz, D. and Yor, M. (1991): *Continuous Martingales and Brownian Motion*. Springer-Verlag, Berlin.
- [20] Shete, S. and Sriram, T.N. (2003): A note on estimation in multitype supercritical branching processes with immigration. *Indian J. Statist.* **65**, 107-121.
- [21] Venkataraman, K.N. (1982): A time series approach to the study of the simple subcritical Galton-Watson process with immigration. *Adv. Appl. Probab.* **14**, 1-20.
- [22] Watanabe, S. (1969): On two dimensional Markov processes with branching property. *Trans. Amer. Math. Soc.* **136**, 447-466.
- [23] Wei, C.Z. and Winnicki, J. (1989): Some asymptotic results for the branching process with immigration. *Stochastic Process. Appl.* **31**, 261-282.
- [24] Wei, C.Z. and Winnicki, J. (1990): Estimation of the means in the branching process with immigration. *Ann. Statist.* **18**, 1757-1773.